9/12/16

Discrete Random Variables:

Ω: finite or countably infinite

$X : \Omega \to \mathbb{R}$

$\#\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$

Continuous Random Variables:

An r.v. X defined on an uncountably infintie sample space is called continuous if $P(X = x) = 0 \quad \forall \ x \in \mathbb{R}$

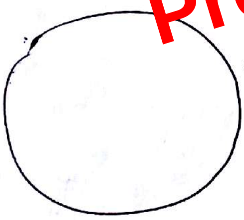
$\Omega - [0,1]$, cannot ennumerate in a list

all infinitely long sequences of 0's & 1's

Ex: $X \sim$ uniform $(0,1)$ if $\forall \quad 0 \le a < b < 1$

$P(X \in (a,b]) = b - a$

Statistics:



population of individuals

N: size of population

N_p: # democrats

$N(1-p)$: # republicans

p: parameter

$p = \frac{N_D}{N}$ ← # democrats

Statisticians provide an estimate of p based on data → representative subset of the population

Data are random - generated through a random experiment.

Gathering data: SRSWR or SRSWOR

SRSWR ↑ with replacement

SRSWOR ↑ without replacement

Y: response variable

$(x_1, \ldots, x_m)$: covariates (the other variable)

Y: phenotype / response to certain kinds of treatment

Genotype: genetic information forms some of the covariates, environmental / cultural factors

Bivariate Density Function:

Let (X,Y) be a continuous random vector

i.e. $P(X=x, Y=y) = 0$

(X,Y) has a density function if there exists a non-negative function

$f: \mathbb{R}^2 \longrightarrow [0, \infty)$ such that for any $A \subseteq \mathbb{R}^2$,

$P((X,Y) \in A) = \int_A f(x,y)\,dx\,dy$

$$= \int_{-\infty}^{\infty} \frac{1}{\pi} I\left\{ y \in \left(-\sqrt{1-x^2}, \sqrt{1-x^2}\right)\right\} dy$$

$$= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{y}{\pi}\Bigg]_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}$$

(not uniform) $$= \frac{2\sqrt{1-x^2}}{\pi}, \quad x \in [-1,1]$$

$$f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi} \quad I\left(y \in [-1,1]\right)$$

conditionals:

$$f_{Y|X=x}(y) = \frac{f(x,y)}{f_X(x)} = \frac{\frac{1}{\pi} I(x^2+y^2 \le 1)}{\frac{2}{\pi}\sqrt{1-x^2}}$$

$$= \frac{1}{2\sqrt{1-x^2}} I\left(y \in \left(-\sqrt{1-x^2}, \sqrt{1-x^2}\right)\right)$$

(uniform) $-1 \le x \le 1$

Given r.v. X

2 important functionals f(x) satisfies are

E(x) & var(x)

same as $\sum y_j P_Y(y_j)$

similar results if X continuous

X has pdf f_X

$Y = g(X)$

$E(Y) = \int g(x) f_X(x)\,dx$

Joint Distributions:

$(X_1, X_2, \dots, X_n)$ is a continuous random vector with pdf $f(x_1, x_2, \dots, x_n)$

$P(X_1 \in (a_1, b_1), X_2 \in (a_2, b_2), \dots, X_n \in (a_n, b_n))$

$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, \dots, x_n)\, dx_n \dots dx_1$$

Think about $g(x_1, x_2, \dots, x_n)$

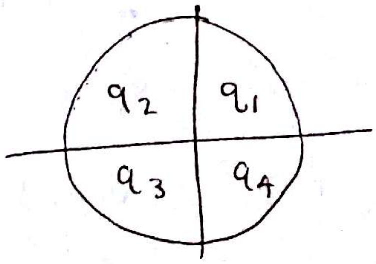
$$E(g(X_1, \dots, X_n)) = \int g(x_1, \dots, x_n) f(x_1, \dots, x_n)\, dx_1 \dots dx_n$$

(X, Y)

$f(x,y) = \frac{1}{\pi} I\{x^2 + y^2 \le 1\}$ $\qquad g(x,y) = xy$

$$E(g(X,Y)) = \int_{x^2+y^2 \le 1} xy \frac{1}{\pi} I\{x^2+y^2 \le 1\}\, dx\, dy$$

$$= \int_{q_1} xy\,dx\,dy + \int_{q_2} xy\,dx\,dy + \int_{q_3} xy\,dx\,dy + \int_{q_4} xy\,dx\,dy = 0$$



We define $E[Y|X]$, the random variable as $\xi(X)$

← not the same as scalar ξ from before

Properties of $E[Y|X]$:

(1) $E(E(Y|X)) = E(Y)$

(2) $var(Y) = E(var(Y|X)) + var(E(Y|X))$

$$var(Y|X=x) = \begin{cases} \sum (y - E(Y|X=x))^2 \, p_{Y|X=x}(y) & \text{discrete} \\ \int (y - \xi(x))^2 f_{Y|X=x}(y)\, dy & \text{continuous} \end{cases}$$

$\xi(x) = E[Y|X=x]$

prediction: $E[(Y - \Psi(X))^2]$

$$= E[(Y - E(Y|X) + E(Y|X) - \Psi(X))^2]$$

$$= E[(Y - E(Y|X))^2 + (E(Y|X) - \Psi(X))^2] +$$

$$2E[\underbrace{(Y - \xi(X))(\xi(X) - \Psi(X))}_{W}]$$

(X, W) $\quad E(W) = E[E(W|X)]$

$$E[W|X] = E[(Y - \xi(X))(\xi(X) - \Psi(X))|X]$$

$$= (\xi(X) - \Psi(X))\, E[(Y - E[Y|X]) | X]$$

$$= (\xi(X) - \Psi(X))(E[Y|X] - E[Y|X]) = 0$$

$$= E[(Y - E[Y|X])^2] + E[E(Y|X) - \Psi(X)]^2$$

The best $\Psi(X)$ is $E[Y|X]$